

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Tinjauan pustaka pada penelitian-penelitian yang dilakukan sebelumnya dalam mendukung penelitian yang sedang dilakukan. Berikut ini adalah penelitian yang telah dilakukan sebelumnya terkait dengan penelitian yang akan dilakukan oleh penulis :

Tabel 2. 1 Tinjauan Pustaka

No	Penulis, Tahun	Judul	Metode
1	Jeffri Prayitno Bangkit Saputra dan Racidon P. Bernarte, 2022	Algoritma Naïve Bayes dalam Memprediksi Penyebaran Omicron Varian Covid-19 di Indonesia: Implementasi dan Analisis	<i>Naïve Bayes Classifier</i>
2	Alvina Felicia Watratan, Arwini Puspita dan Dikwan Moeis, 2020	Implementasi Algoritma <i>Naïve Bayes</i> untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia	<i>Naïve Bayes Classifier</i>
3	Andi Akram Nur Risal, Nur Inayah Yusuf, Andi Baso Kaswar, dan Fathiah Adiba, 2021	Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naïve Bayes	<i>Naïve Bayes Classifier</i>
4	Dedi Darwis, Nery Siskawati, Zaenal Abidin, 2020	Penerapan Algoritma <i>Naive Bayes</i> Untuk Analisis Sentimen Review Data Twitter BMKG Nasional	<i>Naïve Bayes</i>
5	Debby Alita, Indah Sari, Auliya Rahman Isnain, Styawati 2021	Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa	<i>Naive Bayes Classifier</i>
6	Dedi Darwis, Eka Shintya Pratiwi, Ferico Octaviansyah Pasaribu 2020	Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia	<i>Support Vector Machine</i>

Tabel 2. 1 Tinjauan Pustaka (lanjutan)

No	Penulis, Tahun	Judul	Metode
7	Muhammad Ghudafa Taufik Akbar, Dwi Budi Srisullstiowati, 2021	Analisa Sentimen Efektifitas Vaksin terhadap Varian Covid-19 Omicron Berbasis Leksikon	<i>Support Vector Machine</i>
8	Auliya Rahman Isnain, Adam Indra Saktim Deby Alta, Nurman Satya Marga 2021	Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM	<i>Support Vector Machine</i>
9	Styawati, Nirwana Hendrastuty, Auliya Rahman Isnain, Ari Yanti Ramadhani 2021	Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode <i>Support Vector Machine</i>	<i>Support Vector Machine</i>
10	Fikri Aldi Nugraha, Nisa Hanum Harani, Roni Habibi dan Jl. Nuraini Siti Fatonah, 2020	Analisis Sentimen Social Distancing dan Physical Distancing Media Sosial Twiter Menggunakan Algoritma <i>Recurrent Neural Network (RNN)</i>	<i>Recurrent Neural Network (RNN)</i>

Literatur 1 (Saputra & Bernarte, 2022) dengan judul “Algoritma Naive Bayes dalam Memprediksi Penyebaran Omicron Varian Covid-19 di Indonesia Implementasi dan Analisis” dengan permasalahan ditemukan kesulitan yang terkait dengan tingkat penyebaran covid-19 yang terjadi khususnya di Indonesia, dengan isu-isu saat ini kemudian dinilai untuk memutuskan bagaimana mengatasinay dan menentukan luasnya masalah yang perlu diselidiki lebih lanjut. Proporsi instan yang diklasifikasikan dengan benar adalah 48,4848%, sedangkan persentase instance tang tidak diklasifikasikan adalah 51,5152%. Sebanyak 33 data kasus covid-19 per provinsi berhasil dikategorikan dengan tepat, dengan 16 data kasus covid-19 per Provinsi berhasil dikategorikan dengan 16 data terklasifikasikan dengan benar dan

mendapatkan nilai akurasi 46,4252% dan sebanyak 17 data kasus covid-19 per Provinsi salah diklasifikasikan dengan mendapatkan nilai akurasi sebesar 46,4252%.

Literatur 2 (Watratana, B, & Moeis, 2020) dengan judul “Implementasi Algoritma *Naïve Bayes* untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia” dengan permasalahan cirus covid-19 dapat menyebabkan penyakit pada burung dan mamalia termasuk manusia. Pada manusia, corona virus menyebabkan infeksi saluran pernapasan yang umumnya ringan, seperti pilek, beberapa bentuk penyakit seperti: SARS, MERS dan Covid-19. Hasil penelitian menunjukkan bahwa 16 data dari 33 data yang di uji dalam kasus covid-19 per provinsi dengan keakuratan sebesar 48,4848%, dimana dari 33 data yang di uji dalam kasus Covid-19 per provinsi yang diuji terdapat 16 data yang berhasil diklasifikasikan dengan benar.

Literatur 3 (Risa, Yusuf, & Kaswar, 2021) dengan judul “Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma *Naïve Bayes*” dengan permasalahan terdapat lebih dari 100 kandidat vaksin covid-19 yang dikembangkan dan beberapa masih dalam tahap uji coba dan belum terdapat vaksin yang benar-benar efektif. Hasil penelitian menggunakan Algoritma *Naïve Bayes* mendapatkan akurasi sebesar 91% dimana dari 23 data latih, 2 diantaranya tidak diklasifikasikan secara tepat.

Literatur 4 (Darwis, Siskawati, et al., 2020) dengan judul “Penerapan Algoritma *Naïve Bayes* Untuk Analisis Sentimen Review Data Twitter BMKG Nasional” dimana Twittwe menjadi salah satu tempat yang dimanfaatkan untuk menyampaikan informasi yang diberikan BMKG Nasional. Hasil penelitian

menunjukkan Akurasi Algoritma *Naive Bayes* dalam melakukan klasifikasi sebesar 69,97%.

Literatur 5 (Alita et al., 2021) dengan judul “Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa” penelitian ini dilakukan untuk proses penentuan beasiswa sesuai dengan kriteria sehingga menjadi pembantu pengambil keputusan penerima beasiswa menggunakan metode *Naive Bayes Classifier*.

Literatur 6 dengan judul “ Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia” (Darwis, Pratiwi, et al., 2020) masyarakat memanfaatkan Twitter sebagai media penyalur opini terhadap kinerja KPK RI. Klasifikasi dilakukan dengan menggunakan metode *SVM* dengan hasil evaluasi akurasi sebesar 82%.

Literatur 7 dengan judul Analisa Sentimen Efektifitas Vaksin terhadap Varian Covid-19 Omicron Berbasis Leksikon (Akbar & Srisusilowati, 2021) dengan permasalahan peningkatan penyebaran Covid-19 tentang munculnya varian baru covid-19 yaitu omicron. Dengan menggunakan Metode *Support Vector Machine* penelitian ini adalah dataset yang diambil dengan cara *crawling* komentar youtube kanal berita berbahasa ingris kemudian dilakukan pelabelan teks berbasis leksikon menggunakan TextBlob dan Vader ke dalam komentar positif, negatif, dan netral. Terdapat perbedaan jumlah komentar positif, negatif, dan netral dengan kedua metode pelabelan text tersebut namun tidak terlalu jauh. Akurasi penerapan SVM terhadap dataset yang dilabeli dengan metode VADER mengungguli pelabelan komentar dengan metode TextBlob.

Literatur 8 (Isnain et al., 2021) dengan judul “ Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM” dengan permasalahan pendapat masyarakat tentang covid-19 di Indonesia dan dengan adanya virus corona ini banyak korban yang meninggal sehingga membuat masyarakat mengeluarkan opini mereka akan munculnya Virus Covid-19 dan Kebijakan Pemerintah dalam menanganinya. Penggunaan metode *Support Vector Machine* menunjukkan Akurasi sebesar 74%, Presisi sebesar 75%, *Recall* sebesar 92%, dan *F1 Score* sebesar 83%.

Literatur 9 (Styawati et al., 2021) dengan judul “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode *Support Vector Machine*” dengan permasalahan maraknya pembahasan mengenai Kartu Prakerja sebagai salah satu Program Pemerintah pada tahun 2020 dalam mengatasi pengangguran dan korban PHK. Hasil evaluasi yang dilakukan menunjukkan akurasi kernel linear sebesar 98.87%.

Literatur 10 dengan judul “Pengelompokkan Sentimen Pada Twitter Tentang Pendapat Masyarakat Karantina Selama Pandemi Covid-19 Menggunakan Metode DBSCAN, (Manis et al., 2021). dengan permasalahan dilakukan karantina untuk membatasi terjadinya penyebaran penyakit sehingga menimbulkan pro dan kontra dan membuat masyarakat mengungkapkan segala pendapat dan kritikan melalui twitter. Berdasarkan hasil pengujian dilakukan dengan data sebanyak 200 data mendapatkan nilai *silhouette coefficient* terbaik yaitu 0,32 pada nilai *epsilon* sebesar 20 dan *minPts* sebesar 15, sedangkan nilai *davies bouldin index* terbaik yaitu 0,10 pada nilai *epsilon* sebesar 15 dan *minPts* sebesar 15. Penelitian ini

mendapatkan hasil analisis lebih banyak pada sentimen netral yang berarti bahwa masyarakat berpendapat netral terhadap karantina selama pandemi Covid-19.

2.2 Keaslian Penelitian

Berdasarkan tinjauan pustaka yang telah dilakukan oleh penulis, maka perbedaan antara penelitian terdahulu dan penelitian yang dilakukan adalah sebagai berikut:

1. Data set yang digunakan merupakan *tweet* yang berisi komentar masyarakat Indonesia mengenai masuknya varian omicron ke Indonesia.
2. Data set yang digunakan diperoleh dengan cara *crawling* menggunakan Twitter API.
3. Penerapan dan visualisasi algoritma *Naïve Bayes Classifier* menggunakan *software* Rapid Miner.

2.3 Data Mining

Menurut *Gartner Group* data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2006).

Menurut (Santosa & Umam, 2018), data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. sedangkan menurut (Khomarudin, 2016) data mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data, informasi dan knowledge yang didapat tersebut dapat digunakan pada banyak bidang seperti manajemen bisnis, pendidikan, kesehatan dan sebagainya.

2.3.1 Fungsi Data Mining

Menurut (Larose, 2005) dalam bukunya yang berjudul “*Discovering knowledge in data : an introduction to data mining*”, fungsi data mining dibagi menjadi:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numeric dari pada ke arah kategori. Model di bangun dengan *record* lengkap menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai variabel target di buat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan tehnik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat di pisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

5. Pengklasteran

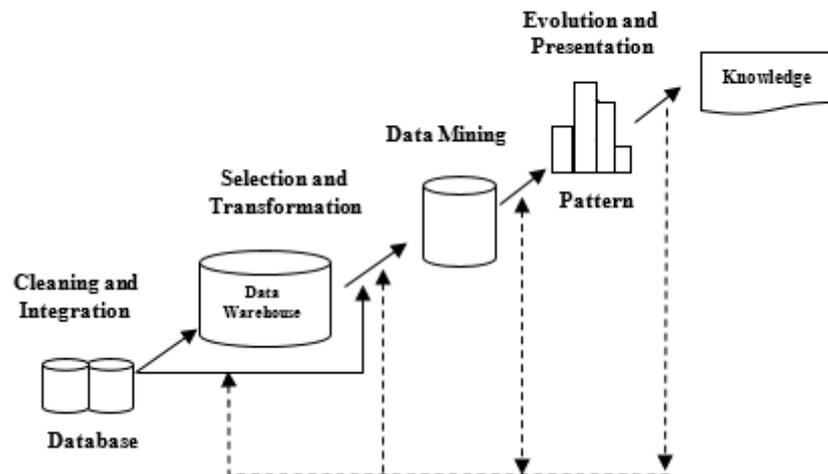
Pengklasteran merupakan penglompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam klaster lain.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analysis).

2.3.2 Tahapan Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 2. berikut tahap-tahap yang ada pada data mining. Proses data mining terbagi dalam enam tahapan (Abdillah et al., 2016), yaitu:



Gambar 2. 1 Tahapan Data Mining (Abdillah et al., 2016)

Adapun penjelasan dari tahapan data mining pada ilustrasi gambar diatas adalah sebagai berikut:

1. Pembersihan Data (*Data Cleaning*)

Proses pembersihan data merupakan kegiatan membuang data yang sama (duplikasi), memeriksa data yang tidak konsisten, dan memperbaiki data yang salah.

2. Integrasi Data (*Data Integration*)

Proses integrasi data merupakan kegiatan menggabungkan data yang berasal dari berbagai sumber.

3. Seleksi Data (*Data Selection*)

Proses seleksi data ini akan dipilih data-data yang hendak digunakan dalam proses sesuai dengan kebutuhan analisis..

4. Transformasi Data (*Data Transformation*)

Pada tahap ini, data yang telah dipilih akan diubah ke dalam format yang telah ditentukan untuk kemudian di proses dan siap ditambah.

5. Proses Mining

Proses mining dilakukan dengan menerapkan metode data mining ke sistem sehingga diperoleh informasi yang diinginkan dari data-data tersebut.

6. Presentasi Pengetahuan (*Knowledge Presentation*)

Tahap ini merupakan tahap dimana hasil dari proses mining diterjemahkan kedalam bentuk informasi yang dapat dipahami oleh pihak yang membutuhkan.

2.4 Text Mining

Menurut (Aditya, 2015), *Text mining* adalah proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut.

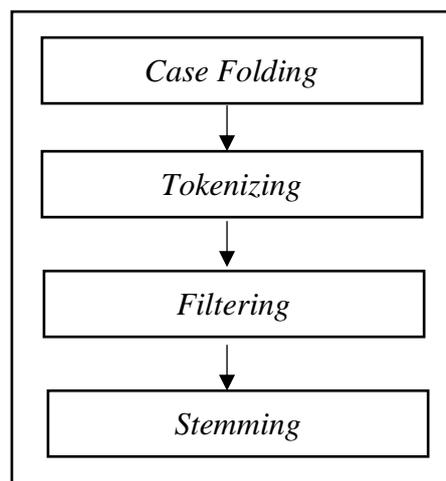
Text mining dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian dan menganalisa *unstructured text* dalam jumlah besar. Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain seperti data mining, *information retrieval*, *statistic* dan matematika, *machine learning*, *linguistic*, *natural language processing*, dan *visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan penyimpanan teks, *processing* akan konten teks, pengumpulan data *statistic* dan *indexing* dan analisa konten.

Text mining memiliki permasalahan yang sama seperti data mining yaitu jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah, data *noise*. Terdapat perbedaan data yang digunakan antara data mining dan *text*

mining. Data mining menggunakan data *structured data*, sedangkan *text mining* umumnya menggunakan data *unstructured data*, atau minimal *semistructured*.

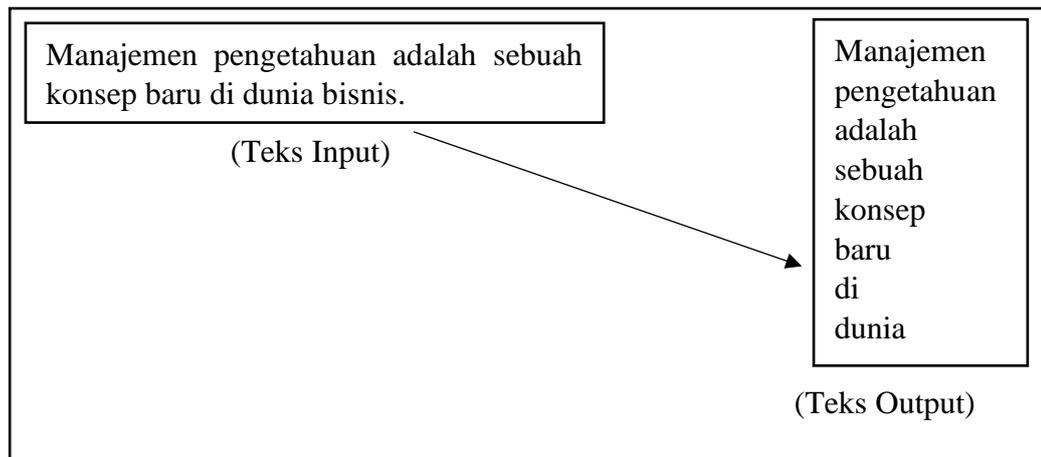
Pada umumnya teks mining memiliki beberapa karakteristik diantaranya yaitu dimensi yang tinggi, terhadap *noise* pada data, dan terdapat struktur teks yang tidak baik. Dalam mempelajari struktur data teks harus menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen, sebelum menentukan fitur-fitur yang mewakili, diperlukan tahap *preprocessing* yang dilakukan secara umum dalam teks mining pada dokumen, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming* (Raymond, 2006), seperti terlihat pada gambar

2.1



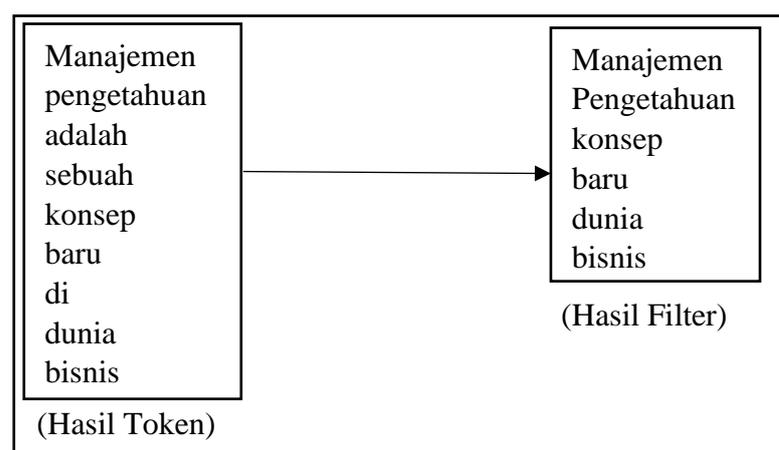
Gambar 2. 2 Proses Text Mining

Case folding yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf `a` sampai dengan `z` yang diterima. Karakter selain huruf dihilangkan dan dianggap deimeter. proses *folding* seperti gambar 2.2.



Gambar 2. 3 Proses Folding

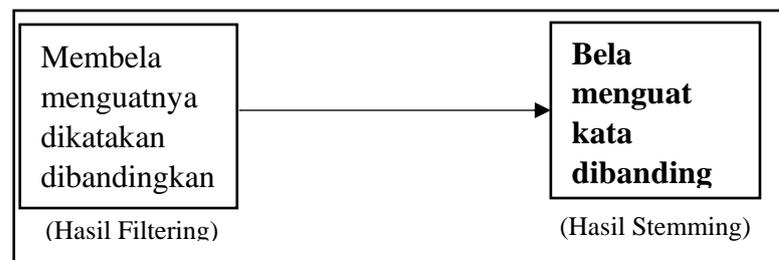
Tahap *Tokenizing* atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya, sedangkan tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil *term*. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-word*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan seterusnya. Proses *tokenizing* dan *filtering* seperti gambar 2.3.



Gambar 2. 4 Proses Tokenizing dan Filter

Tahap *stemming* adalah tahap mencari *root* kata dari tiap kata hasil *filtering*. Pada tahap *stemming* dilakukan proses pengambilan berbagai bentukan kata

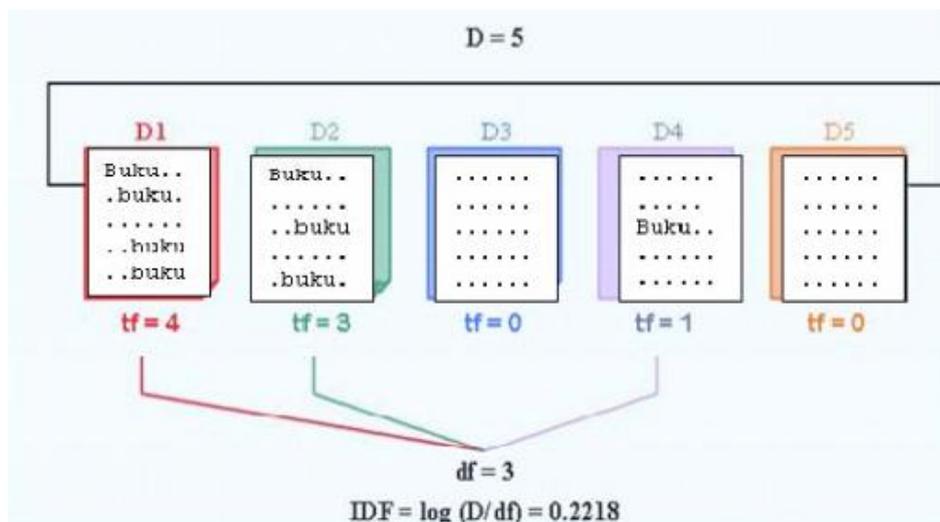
kedalam suatu representasi yang sama. Tahap ini kebanyakan dipakai untuk teks berbahasa inggris dan lebih sulit diterapkan pada teks berbahasa indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen. Proses tahapan *stemming* pada teks berbahasa Indonesia seperti pada gambar



Gambar 2. 5 Proses *Stemming*

2.4.1 Algoritma *TF-IDF*

Algoritma *TF-IDF* merupakan algoritma pembobotan data berbentuk teks menjadi *numeric* atau *binary* sehingga data dapat diterapkan kedalam algoritma *data mining*. Algoritma ini merupakan jenis klasifikasi yang didasarkan oleh relevansi umpan balik dan diusulkan oleh (Rocchio, n.d.). Gambar 2.6 berikut ini merupakan ilustrasi dari Algoritma *TF-IDF*.



Gambar 2. 6 Ilustrasi Algoritma *TF-IDF*

Keterangan:

D1, D2, D3, D4, D5 = Dokumen

D = Total Dokumen

Tf = Banyaknya kata yang dicari pada sebuah dokumen

Df = Banyak dokumen yang mengandung kata yang dicari

Terdapat 3 tahap dalam menggunakan algoritma ini, dimana tahapannya adalah sebagai berikut:

1) *Term Frequent*

Tf adalah frekuensi dari suatu kata yang muncul dalam satu dokumen.

2) *Inverse Document Frequency*

IDF adalah nilai *invers* dari dokumen yang mengandung kata tersebut.

Berikut ini merupakan persamaan untuk menghitung nilai IDF :

$$IDF(t) = \log\left(\frac{n}{df(t)}\right) \quad (2.1)$$

Dimana: n = total dokumen

$df(t)$ = jumlah dokumen yang mengandung kata t

3) *Weighting TF-IDF (Pembobotan)*

Untuk menemukan nilai bobot dari data dilakukan dengan cara mengalikan nilai TF dan IDF . Berikut ini merupakan persamaan dari pembobotan $TF-IDF$:

$$W(dt) = TF(d, t) \times IDF(t) \quad (2.2)$$

2.4.2 Analisis Sentimen

Menurut (Mujilahwati, 2016), Analisis Sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas, dimana entitas tersebut dapat mewakili

individu. Menurut (I et al., 2012), analisis Sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif. Covid-19 adalah penyakit jenis baru yang.

2.5 *Naïve Bayes Classifier*

Algoritma naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger, 2007). *Naive Bayes Classifier* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas, (Patil & Sherekar, 2013). Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Persamaan Algoritma Naive Bayes Classifier adalah, (Bustami, 2013)

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2.4)$$

Di mana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk mengetahui nilai dari data yang akan diprediksi akan dilakukan penghitungan pemetaan data menggunakan $Vmap$, perhitungan nilai $Vmap$ dilakukan dengan menggunakan persamaan sebagai berikut:

$$Vmap = V_j\{Positif, Negatif\}P(V_j) \prod P(X_i | V_j) \quad (2.5)$$

2.6 Covid-19

Corona virus merupakan virus yang dapat menyebabkan gejala ringan sampai berat. setidaknya ada 2 jenis corona virus yang diketahui dapat menyebabkan penyakit dengan gejala berat seperti *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). Covid-19 adalah penyakit yang belum pernah diidentifikasi sebelumnya, virus penyebab Covid-19 ini disebut dengan Sars-Cov-2. SARS dibawa atau ditransmisikan oleh luwak ke manusia dan MERS dari unta ke manusia, (Organization, 2020).

2.7 Twitter dan Twitter API

2.7.1 Pengertian Twitter

Menurut (Ahuja & Dubey, 2017), Twitter merupakan *platform* media sosial yang digunakan orang-orang diseluruh dunia sebagai tempat yang tepat untuk menyampaikan pendapat mereka. Menurut (Hasan, 2017), Twitter adalah suatu jejaring sosial yang membatasi penggunaannya untuk mengirim sebuah tweet dengan 140 karakter tidak lebih. Twitter merupakan layanan jejaring sosial yang berguna untuk menghubungkan antara pengguna satu dengan pengguna lainnya.

2.7.2 Struktur Data Twitter

Pesan *twitter* memiliki banyak ragam struktur data.(A. Go & L. Huang, 2009), menjabarkan karakteristik *twitter* sebagai berikut :

1. Pada pesan *twitter*, setiap *tweet* hanya berisi panjang maksimal 140 karakter, dengan nilai rata-rata panjang setiap *tweet* dimana diketahui rata-rata *tweet* adalah 14 kata atau 78 karakter.
2. Data *twitter* dapat bersumber dari beberapa tempat. Dengan *twitter* API data dengan mudah didapat.
3. Pengguna *twitter* dapat dengan mudah menggunakan media apapun untuk menulis dan mengirimkan pesan mereka, termasuk penggunaan media ponsel. Kemunculan kesalahan penulisan ataupun penggunaan bahasa asing jauh lebih tinggi. Terdapat ragam topik didalamnya.

Menurut (Davido & Tsur, 2010), menyimpulkan bahwa sebuah *tweet* biasanya mengandung alamat URL, alamat pengguna twitter yang disebut *username* (@+*username*), atau konten *tag* yang disebut *hashtag* (#), dan *emoticon*. Penggunaan *hashtag* dan *emoticon* dianggap juga dapat tidak mewakili dari sentimen dalam sebuah *tweet* (A. Go & L. Huang, 2009)

2.7.3 Pengertian Twitter API

Twitter Application Programming Interface (API) merupakan fungsi-fungsi untuk menggantikan bahasa yang digunakan dalam *system calls* dengan bahasa yang lebih terstruktur dan mudah dimengerti oleh programmer. Fungsi yang dibuat dengan menggunakan API kemudian akan memanggil *system calls* sesuai dengan sistem operasinya. Tidak menutup kemungkinan nama dari *system calls* sama

dengan nama di API (Arifin & Suryatul). API twitter terdiri dari 3 (tiga) bagian yaitu:

1. *Search API*

Search API dirancang untuk memudahkan *user* dalam mengelola *query search* di konten twitter. *User* dapat menggunakannya untuk mencari *tweet* berdasarkan *keyword* khusus atau mencari *tweet* lebih spesifik berdasarkan *username twitter*. *Search API* menyediakan akses pada *Data Trending Topic*.

2. *Representational State Transfer (REST) API*

Representational State Transfer (REST) API memperbolehkan *Developer* untuk mengakses inti dari twitter seperti *timeline*, *status update*, dan informasi pengguna. *Representational State Transfer (REST) API* digunakan dalam membangun sebuah aplikasi twitter yang kompleks yang memerlukan inti dari *twitter*.

3. *Streaming API*

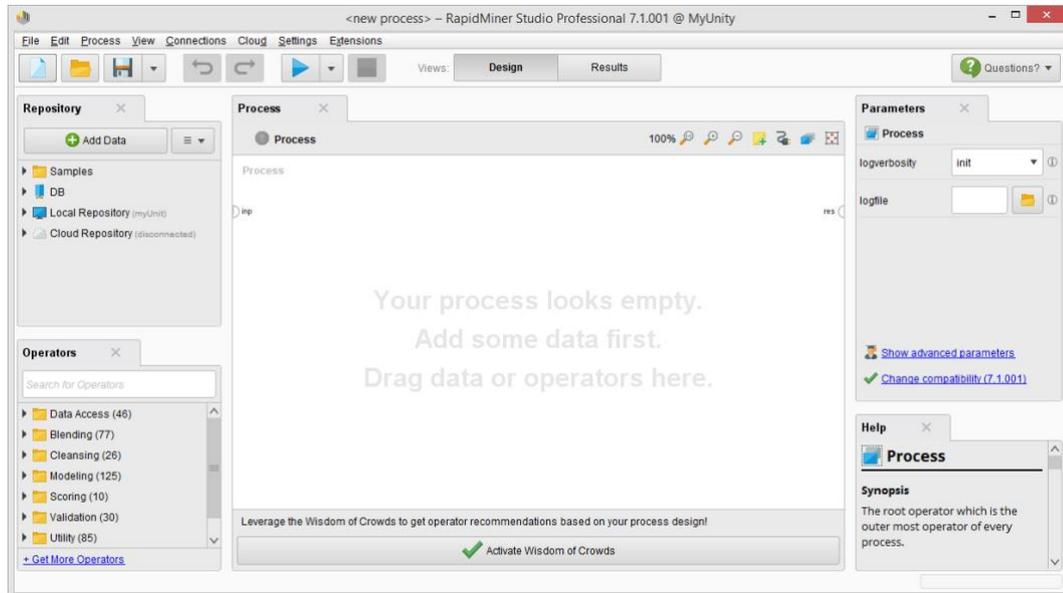
Streaming API memperbolehkan *developer* untuk kebutuhan yang lebih intensif seperti melakukan penelitian dan analisis data. *Streaming API* menghasilkan aplikasi yang dapat mengetahui statistik *status update*, *follower*, dan lain sebagainya.

2.8 Rapid Miner

Penelitian ini menggunakan *tools Rapid Miner 7.1* untuk menerapkan *association rule* terhadap data penelitian yang digunakan. Rapid Miner adalah *software* yang bersifat *open source* untuk melakukan analisis terhadap *data mining*, *text mining*, dan analisis prediksi, (L. Elvitaria & M. Havenda, 2017). Rapid Miner merupakan sebuah solusi untuk melakukan analisis terhadap data mining dengan

menggunakan teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik, (C, 2013) .

Tampilan dari Rapid Miner 7.1 dapat dilihat pada gambar 2.6.



Gambar 2. 7 Tampilan Rapid Miner